Physical principles of building protein megacomplexes in a crowded milieu

Jiayi Wang [1], Jules Nde [2], Andrei G. Gasic,[3] Jacob Haseley,[1] and Margaret S. Cheung [1,4,*]

[1]*Department of Physics, University of Washington, Seattle, Washington, USA*

[2]*Department of Cancer Biology, University of Kansas Medical Center, Kansas City, Kansas, USA*

[3]*Geminus.ai, Inc., Houston, Texas, USA*

[4]*Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, Washington, USA*

(Received 15 February 2026; accepted 8 August 2026; published 15 September 2026)

Multiple phenotypic protein expressions arise from one genome represent variations in the protein relative abundance and their stoichiometry. A lack of definite compositional parts challenges the modeling of protein megacomplexes and cellular architectures. Despite the advance in protein structural predictions with Artificial Intelligence (AI), the mechanism of protein interactions and the emergence of megacomplexes they assemble remains unclear. Here, we present a statistical-physics framework of grand canonical ensemble to explore the protein interactions that drive the emergent assembly of a megacomplex using the observational mass spectrometry datasets including protein relative abundance and the cross-linked connections. Using chromatin remodeler megacomplex, INO80, as an example, we discovered a class of "divergent" protein that plays a critical role in orchestrating the assembly beyond nearest neighbors, dependent on the excluded volumes exerted by others. With the constraints of the excluded volumes by varying the volume fraction of INO80 subunits, these divergent subunits orchestrate and form clusters with selective components growing into configurationally distinct architectures. We propose a machinery view for the INO80 chromatin remodeler complex where each loosely associated subunits can be occasionally recruited for parts as attachment into a core assembly driven by excluded volumes. Our computational framework provides a mechanistic insight into taking the volume fractions as necessary physicochemical variables representing cell states to remodel the configurations of protein megacomplexes with structurally loose modules.

DOI: 10.1103/b8cy-vfy5

I. INTRODUCTION

There has been a rapidly growing interest in understanding why many cellular phenotypes arise from one genome in the content of environmental gradients enabled by high-throughput experimentations, imaging, and advance analytic tools guided by Artificial Intelligence (AI) [1]. Given that genomic information and derived knowledge is abundantly available, we gained insight about biological networks from gene expression data [2–4], single-cell RNA sequencing [5], and protein sequences and structures [6]. However, the multiple phenotypic expressions at the proteomic level involving variations in protein interactions and their abundance representing a cell state [7] remain poorly understood [Fig. 1(a)] [8,9]. Protein assemblies dynamically adapt to a cell state [10,11] by associating into biomolecular condensates [12–16] or megacomplexes [17–20]. Estimated by the length scale of interactions and features of the two spatially distinctive structures [21]: The former denotes a dense phase in the cytoplasm formed by liquid-liquid phase separation in a subcellular scale [12] and the latter denotes structurally hierarchical protein aggregates arising from higher-order interactions [17] in a molecular scale. Both are outstanding examples that manifest a growing complexity from genotype to phenotype expressions shaped by environmental influences. Despite the advance in protein structural predictions with AI, the mechanism of protein interactions and the emergence of megacomplexes they assemble remains unclear.

To address this gap, several computational and theoretical frameworks have been developed [12–16,22–24] to investigate the protein interactions that drive macromolecular assemblies involving heterotypical constituents. One approach is the scaffold-client framework, in which proteins interact with more partners (i.e., higher valency) are considered scaffolds to recruit the clients interacting with few partners (i.e., lower valency) [25,26]. While there are extended models to include more comprehensive interactions between scaffolds and clients, this framework is still under the assumption of a fixed protein level [27,28], representing a static cell state.

To incorporate the consideration of dynamic cell states that influence the constituents as well as the configurations of protein assemblies, we deployed an open system based on a grand canonical ensemble framework [29] to test a hypothesis that the excluded volumes of subunits limit the configurational search of protein interactions in a reservoirlike

*Contact author: margaret.cheung@pnnl.gov

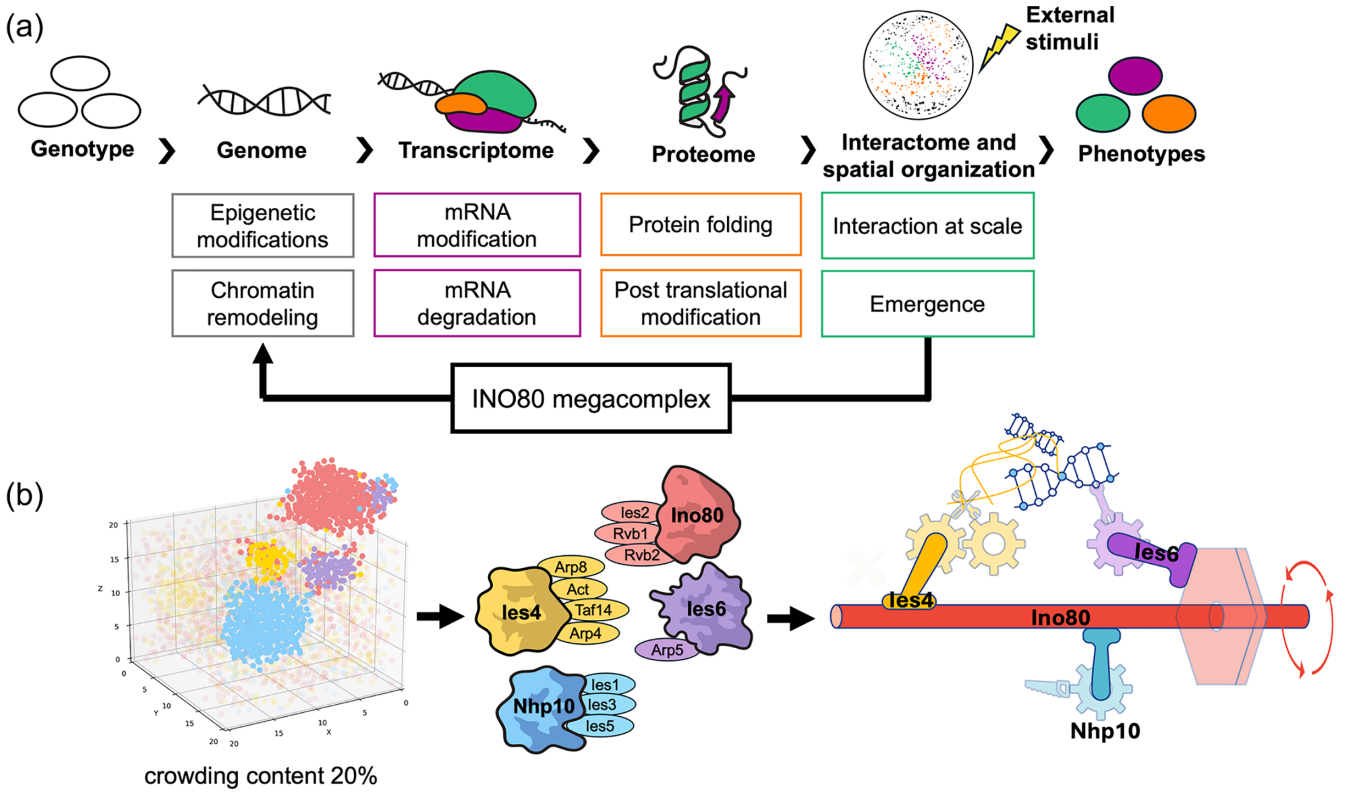

FIG. 1. Protein spatial organization mechanisms as a knowledge gap. (a) Transition from genotype to phenotype, where understanding of proteins' spatial arrangement under external stimuli is critically lacking. The INO80 chromatin remodeling complex, which modifies chromosome topology according to cell state in the yeast metabolic cycle, is chosen as a candidate to gain insight into the protein's spatial organization. (b) The simulated cluster formation at a specific volume fraction and the inferred subunits' hierarchical order in spatial regulation. Red, purple, blue, and yellow represent subunits in the Ino80 core, Arp5, Nhp10, and Arp8 modules, respectively. Cloud and round shapes represent the difference in hierarchy during subunit gathering.

cytoplasm environment, and that drives protein assemblies. We used yeast chromatin remodeler INO80 complex [30–36] as an example for the assembly of a megacomplex with many structurally loose parts in a cell-like environment and they remodels [30,31] upon environmental cues.

We represented a dynamic cell state where a system continuously exchanges components and moves toward equilibrium with the surrounding particle reservoir. We generated structural models by inferring interaction energies and chemical potentials between loose parts from experimental data for INO80. Through thermodynamic perturbations such as variation in particle number and excluded volume [37–40], we teased out the critical components in INO80 that orchestrates interactions beyond nearest neighbors. The excluded volumes contributed to the depleted interactions [41] between modules at high volume fraction, giving rise to distinct configurational state of structural machinery for INO80 under various total concentrations [Fig. 1(b)].

II. METHOD

A. Composition of INO80 complex

Yeast INO80 consists of 15 subunits (see the Supplemental Material [42] for a detailed list of INO80 subunits), which are grouped into four modules (bold represents modules): the Ino80 Core (Ino80, Ies2, Rvb1, Rvb2), Arp8 (Arp8, Arp4, Actin, Ies4, Taf14), Arp5 (Arp5, Ies6), and Nhp10 (Nhp10, Ies1, Ies3, Ies5) modules; each module has its distinct function [32–34,43–46].

B. Inversely determine INO80 subunits' interaction energies and chemical potentials

To model the spatial-temporal process of INO80 complex cluster formation in a grand canonical ensemble [Fig. 1(b)], we coarse grain each protein as a bead of the same size. The following input variables are needed: the relative abundance of subunits, interaction energies between subunit types, and chemical potentials. The relative abundance of INO80 subunits can be measured experimentally. However, the interaction energies and chemical potentials can only be inferred from empirical values using a maximum entropy approach, in which we estimate subunit interaction energies and chemical potentials from protein abundance and interaction data.

For the relative abundance of INO80 subunits, we take data from the mass spectrometry study in Ref. [47]. The total number of particles in the simulation box is then determined using the equation $N_{\text{total}} = L^3\phi/v_0$, where L is the length of the cubic box, ϕ is the volume fraction, and $v_0 = \frac{1}{6}\pi\sigma^3$ is the volume of a single bead with σ being the diameter of the beads. We use a box size of $L = 20\sigma$. The justification of box size to be sufficiently large is provided in Fig. S2 in

the Supplemental Mateial [42]. The total particle number is then distributed into multiple particle types according to their relative abundance.

For interprotein interactions, we infer the interaction energies that could reproduce the empirical observations from the cross-linking experiment. The Lennard-Jones potential $U_{ijab}(\boldsymbol{r}) = \frac{\epsilon_0}{|\boldsymbol{r}_i^a - \boldsymbol{r}_j^b|^{12}} - \frac{\epsilon_{ij}}{|\boldsymbol{r}_i^a - \boldsymbol{r}_j^b|^6}$ is used to model the interaction between two beads of types i and j; a and b are the indices of particle, ranging from 1 to N_i^{exp}, and 1 to N_j^{exp} correspondingly. $\boldsymbol{r}$ is the coordinate vector representing the location of each particle. The repulsive interaction energy ϵ_0 is universal to all beads to account for excluded volume, while the attractive interaction energy ϵ_{ij} depends on bead types.

We first estimate an interprotein contact intensity for each pair of subunit types from the cross-linking experiment in Ref. [35]. Then, using the algorithm developed by Zhang and Wolynes [48] and Morcos [49], we iteratively update the interaction energies ϵ_{ij} until the simulated interprotein contact intensities match the experiment observations to a certain extent. To identify the most representative simulation trajectory of INO80, we also consider the known INO80 modules and screen over the range of simulation trajectories that meet the previous selection criteria. The detailed procedure, mathematical derivation, and the resulting interaction energies are shown in the Supplemental Material [42].

After the interaction energies are determined, we proceed to determine each subunit's chemical potential. To do that, we build on the maximum entropy algorithm developed in Gasic and co-workers [29], and enable the subunit's particle exchange with the reservoir one at a time. For every single subunit, we iteratively update the chemical potential while fixing the abundance for the other 14 subunits, until the particle number of the investigated subunit in the simulation matches the experimental observation:

$$\mu^{k+1} = \mu^k - k_l[\langle N\rangle - \tilde{N}]^k, \tag{1}$$

where μ is the chemical potential of the subunit of inquiry, k is the iteration number, and k_l is an adjustable learning rate. $\langle N\rangle$ is the ensemble average of the subunit's simulated particle number at the current iteration, and $\tilde{N}$ is the empirical abundance. The percentage difference $\frac{\langle N\rangle - \tilde{N}}{\tilde{N}}$ is calculated after each iteration; the iteration is stopped when the percentage difference is below 0.01.

C. Simulation of subunits assembly and analysis

Depending on whether a convergence to empirical abundance can be reached, subunits are classified into two types: convergent and divergent (see Sec. III). To simulate the formation of the INO80 megacomplex, convergent subunits freely conduct particle exchange with the cellular reservoir, while the abundances of divergent subunits are held strictly constant. The INO80 assembly is thus modeled in a grand canonical Monte Carlo simulation where a fixed number of divergent subunits recruit convergent subunits from the reservoir based on their chemical potentials. We then investigate the physics principles underlying the convergent and divergent properties to elucidate their role in subcellular spatial organization through the following steps.

1. Mapping chemical potential's relation with particle number

To understand the fundamental physics behind convergence and divergence behavior, we performed simulations over a range of chemical potentials for a single subunit type, while keeping the particle numbers of the others fixed at the empirical values. We then map the subunit's particle number against chemical potentials. Among the 15 subunits, we focused on those exhibiting divergent behavior in at least one volume fraction and analyzed their profiles across all three volume fractions.

To probe the physical basis of this convergence-divergence phenomenon, we also perturbed the temperature to $T = 0.9$ and $T = 1.1$ at a fixed volume fraction of 0.05 and examined the resulting abundance-chemical potential curves. For volume fraction 0.05, we use 1.5 million Monte Carlo steps. For volume fractions 0.1 and 0.2, we use 2 million Monte Carlo steps and multiple initial conditions when needed.

2. Particle number perturbation

To better understand how each subunit's particle number shapes the assemblage, we performed canonical Monte Carlo simulations using empirical particle numbers for all 15 subunits. Then, each subunit's count was varied by ± 10, 20, 30, or 40, together with a test particle with only volume exclusion and no interaction with others for baseline reference. For low-abundant subunits Ies4 and Ies6, perturbations exceeding their empirical counts were marked "N/A." All simulations were run at reduced temperature $T = 1$, and a volume fraction of $\phi = 0.1$ for the reference trajectory without perturbation. The box size remained the same as the unperturbed conditions. At a volume fraction of 0.1, the total particle count in the box is 1525, so the maximum perturbation of 40 test particles leads to a change of 2.6% in total volume fraction, which is marginal to our analysis.

We ran all simulations for 4 000 000 timesteps, dumping 4000 frames (1 frame per 1000 steps). Analyses used every 10th frame from 2000 to 4000 to sample equilibrated trajectories. We analyzed the shifts in the ensemble averages of network features (see the Supplemental Material [42] for detailed methods and results) and thermodynamic properties for each perturbation.

For thermodynamic properties, we chose two indicators—change in available volume ΔV_a [37,50,51] and potential energy ΔE. Available volume is defined as the total space available for a test particle of a given size to be inserted. Consider inserting a test particle into a system of i particles, we have

$$\begin{aligned} V_{\text{excluded}} &= U_i\left[\tfrac{4}{3}\pi(r_i + r_t)^3\right], \\ V_a &= V_{\text{tot}} - V_{\text{excluded}}, \end{aligned} \tag{2}$$

where V_{excluded} is the excluded volume, and V_{acc} is the available volume for associated or unassociated particles [37]. r_i and r_t are the radius of the existing particle i and test particles, respectively. V_{tot} is the total volume of the system (in this case, the volume of the box).

3. Calculating the effective potential mean force

To further distinguish the system-wide effects of divergent and convergent subunits, we use Ovito [52] to compute the

radial distribution function $g(r)$ between selected subunits and all particles in the simulation box. $g(r)$ captures the effect of chemical potential in the grand canonical ensemble because the particle numbers in the simulation box were derived from the chemical potentials self-consistently. We therefore approximated the corresponding effective potential mean force (PMF) using the relation $W(r) = -k_B T \ln g(r)$, where k_B is Boltzmann's constant and T is temperature.

III. RESULTS

A. A small subset of subunits shows "divergent" characteristics in reservoirlike cytoplasm

We model the cellular impact of relative protein abundance into chemical potential [29] and their interactions from cross-linking mass spectrometry for the INO80 complex (see Sec. II). For each 15 subunits of the INO80 complex, we iteratively adjust the chemical potential until the simulated particle number converges to its experimental abundance while keeping the other 14 subunits' abundance fixed. For most subunits, their convergence is reached within a few iterations (see Fig. S4 in the Supplemental Material [42]). However, for a small subset of subunits, the particle number oscillates around the experimental abundance and are unable to converge toward experimental estimates, regardless of hyperparameters. We call this specific category of subunits "divergent." Others are by comparison "convergent."

To better understand the underlying principles of the newly discovered divergent property of subunits in INO80, we plotted the phase diagram of density versus chemical potential of the subunits that display divergent behavior at a reduced temperature unit, $T = 1$, and the volume fractions of subunits, ϕ equals 0.05, 0.1, and 0.2 in Fig. 2. This range of ϕ represents the total volume fractions of INO80 subunits [39,53]. Density represents normalized abundance and ϕ is defined as the proportion of a system's total volume that is occupied by the INO80 subunits. In Fig. 2(a), the profiles for the divergent subunits are plotted in thick lines. They exhibit a discontinuity in density around a critical chemical potential μ^*. For example, the profile of Ies4 (yellow curve) at $\phi = 0.1$ shows a density discontinuity spanning from roughly 0 to 0.014 at chemical potential $\mu^* = -9.8$, within which the empirical value (indicated by the yellow arrow) falls. Similarly, Ino80 also exhibits this discontinuity in density that encompasses its empirical value (indicated by the red arrow). Both Ies4 and Ino80 display consistent divergence property across all volume fractions.

Interestingly, not all subunits display divergent behavior across the whole range of ϕ. Some subunits, such as Ies6, Arp5, and Nhp10, exhibit emergent divergent property only within a narrow range of volume fraction. For example, Ies6 subunit is convergent at $\phi = 0.05$ but become divergent at $\phi = 0.1$ and $\phi = 0.2$. Arp5 and Nhp10, on the other hand, show critical behavior between divergent and convergent (more in Sec. IV). Arp5 only become divergent at $\phi = 0.1$, and Nhp10 does so at $\phi = 0.2$. We highlight them with dotted lines in Fig. 2(a) for visual guidance.

The discontinuity in the chemical potential versus density curves in Fig. 2(a) of a divergent subunit indicates that the system is highly unstable within that region when connecting the subunit to a particle reservoir. A thermodynamic interpretation of the divergence phenomenon is provided using the framework of spinodal decomposition [inset of Fig. 2(a)], in which the particle numbers of divergent subunits fall within the unstable region of the phase transition (more in Sec. IV).

B. Effective stickiness between divergent subunits and others emerges under high volume fraction

Divergent subunits, with their empirical abundance [pointed by arrows in Fig. 2(a)] being within this discontinuity region, are more suitably modeled with a fixed particle number rather than within a reservoir. We thus simulated the INO80 megacomplex formation using the empirical abundance of divergent subunits and the inferred chemical potentials of the rest across the three volume fractions (see Sec. II). The formed clusters and their compositions are shown in Fig. S6 in the Supplemental Material [42]. Interestingly, we observed that all formed clusters consist of at least one divergent subunit.

The divergent property of a subunit, however, depends on volume fraction, as manifested in Ies6. Given that subunit Ies6 subunit is convergent at $\phi = 0.05$ but become divergent at the volume fractions of 0.1 and 0.2, we further explore its interactions with other subunits at these varying volume fraction by computing the effect PMF. In Fig. 2(b), we derived PMF between Ies6 and all other subunits at the three volume fractions, ϕ of 0.05, 0.1, and 0.2, in order to understand the many-body interactions on the properties of Ies6 subunits (see Sec. II). At $\phi = 0.05$, there is only volume repulsion as PMF > 0 everywhere; at $\phi = 0.1$, an attractive well appears, since PMF < 0 when r, the distance between Ies6 and other subunits, is less than 5σ (the diameter of a subunit). Interestingly, when $\phi = 0.2$, PMF first decrease to below 0 at short r, and then increase above 0 when $6\sigma < r < 9\sigma$, representing a desolvation barrier separating the subunits. The presence of desolvation barrier adds the entropy cost to break apart interactions [54], essentially "strengthening" the depleted [41], sticky effect.

The changes in the profiles of effective PMF with ϕ demonstrates the fascinating phenomena where a subunit's effective interaction with the system is shaped by volume fraction. Take Ies6 in Fig. 2(b) as an example, it becomes divergent at $\phi \geqslant 0.1$, corresponding to the fact that it shows stickiness in the effective PMF as opposed to repulsion due to volume exclusion at $\phi = 0.05$. We observed that this behavior can be generalized to other subunits in Fig. S8 in the Supplemental Material [42], which shows that effective PMF for the divergent subunits have an attractive potential well interacting with all other subunits at $\phi = 0.1$. In contrast, the convergent subunits exhibit only volume exclusion with all other subunits.

C. Divergent subunits orchestrate spatial organization beyond local interactions

To further understand the thermodynamic contribution of divergent subunits to the aggregation of assembled clusters, we analyzed the emergent behavior and thermodynamic properties by quantifying each subunit's contribution to the assembly they form. We calculated the free-energy changes

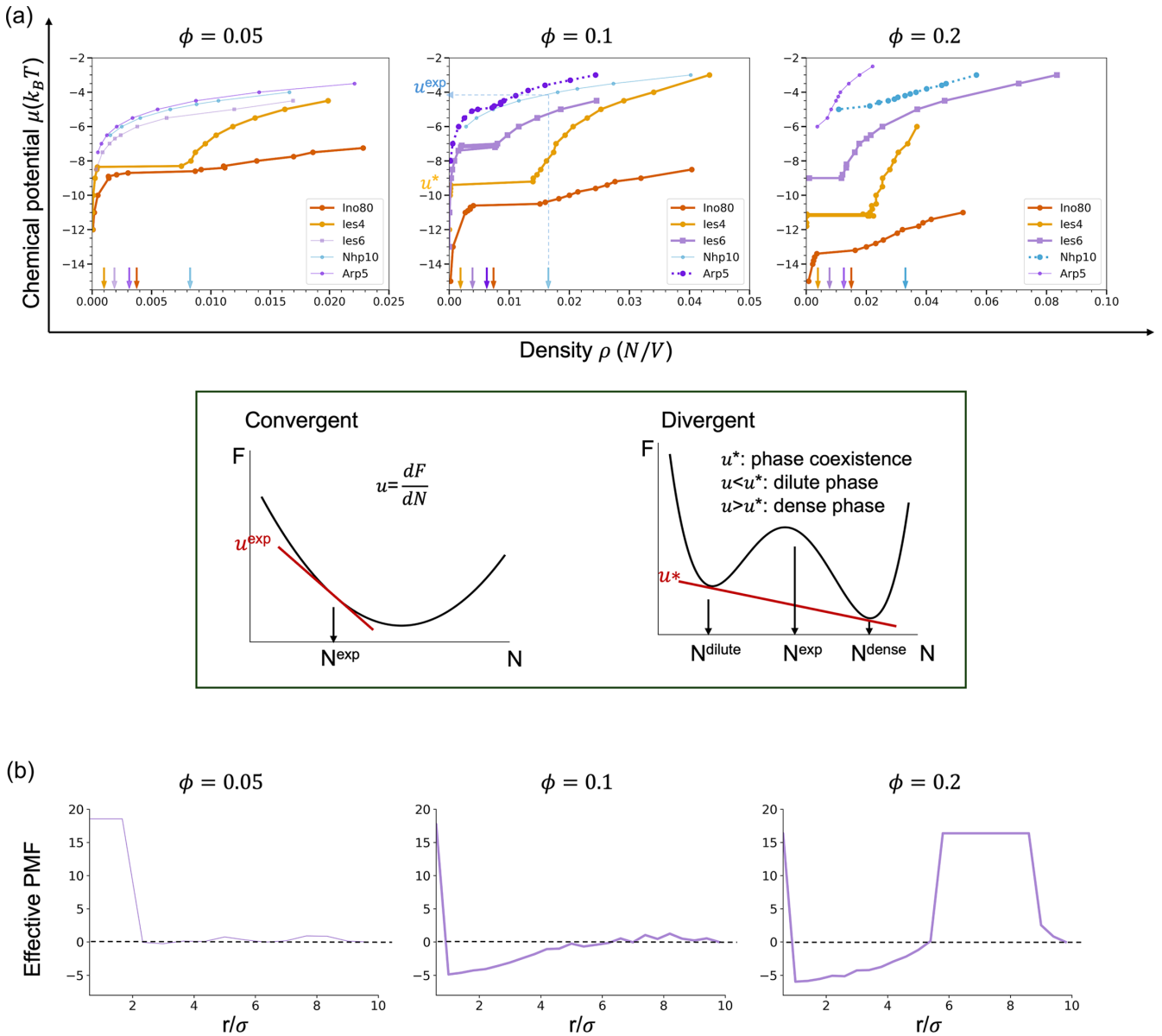

FIG. 2. Chemical potential vs density profile and PMF for subunits showing divergent property at any volume fraction. (a) Chemical potential vs density for subunit Ino80, Ies4, Ies6, Nhp10, and Arp5 at volume fractions $\phi = 0.05,\ 0.1$, and 0.2 and temperature $T = 1$. Colors represent the INO80 module a subunit belongs to. Opaque curves indicate divergent subunits, and transparent curves indicate convergent subunits. Curves with dashed lines represent subunits at the boundary between convergence and divergence. The arrows on the x axis represent each subunit's experimental density. The inset shows schematic diagram of the system's free-energy landscape along the particle number of a single subunit with divergent or convergent property. The idea is adapted from the theory of spinodal decomposition [55]. (b) Effective potential mean force between Ies6 subunit and all particles at volume fractions $\phi = 0.05,\ 0.1$, and 0.2 and temperature $T = 1$.

(ΔG) represented by available volume (ΔV_a) and potential energy (ΔE) resulting from variations in its particle number, dN (see Sec. II). V_a is expressed in the unit of σ^3, and E is in the unit of k_BT, where k_B is the Boltzmann's constant.

To achieve a baseline understanding of particle insertion free energy, ΔG, we measured ΔV_a^0 and ΔE^0 using a test particle with only volume exclusion. We then compared it to the perturbation by inserting particles of each subunit i from INO80 complex as ΔV_a^i and ΔE^i by calculating the difference in perturbation $\Delta\Delta V_a^i = \Delta V_a^i - \Delta V_a^0$ and $\Delta\Delta E^i = \Delta E^i - \Delta E^0$. The result is shown in Fig. 3. From Fig. 3(a), we see that compared to test particles, inserting divergent subunits increases the available volume $(\Delta V_a^i - \Delta V_a^0 > 0$ at $dN > 0)$ and removing them reduces available volume $(\Delta V_a^i - \Delta V_a^0 < 0$ at $dN < 0)$. An insertion of a "sticky" divergent subunit induces nucleation and opens up the available volume for particle insertion by promoting effective attractive interactions with others, significantly lowering the potential energy as shown in Fig. 3(b). They facilitate cluster formation through long range interaction, as manifested in the long tail of an effective potential of mean force. In contrasts, perturbation of particle insertion from convergent subunit remains insignificant, and the impact of convergent subunits on others is simply local.

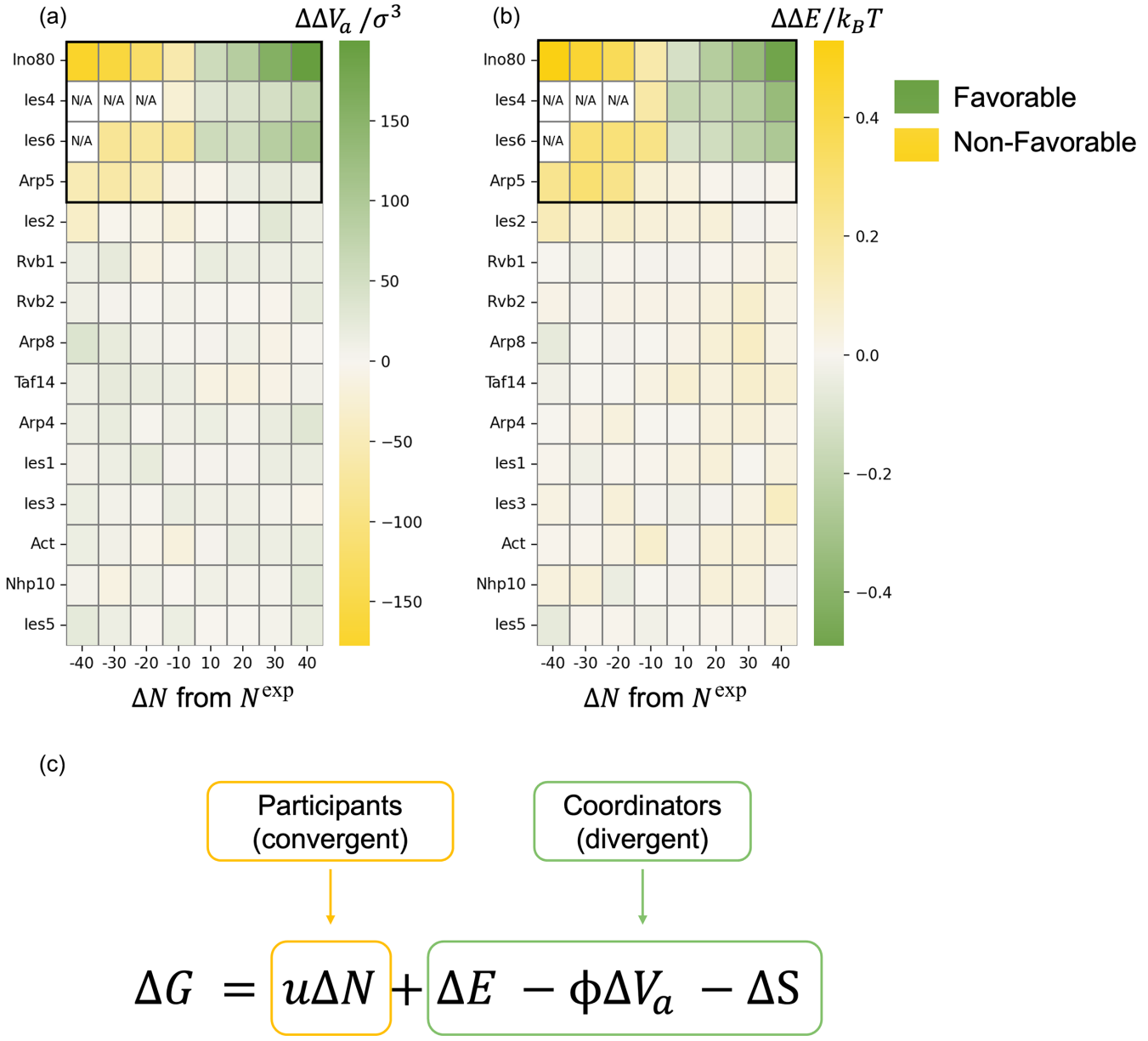

FIG. 3. Free-energy change from particle number variation of each INO80 subunit. (a), (b) Normalized change in available volume $\Delta\Delta V_a^i$ and potential energy $\Delta\Delta E^i$ for each subunit i, respectively, caused by particle number variation at volume fraction $\phi = 0.1$ and temperature $T = 1$. Row represents subunit, and columns correspond to particle number perturbation ΔN. Green indicates thermodynamically favorable responses, while yellow denotes thermodynamically unfavorable ones. Divergent subunits are in the black box. N/A represents an unphysical negative particle number due to low experimental abundance. (c) A theoretical illustration of the divergent and convergent subunits' contribution to the free-energy terms.

To theoretically describe the thermodynamic contributions of divergent and convergent subunits to the overall free-energy change upon particle insertion, we break down into four terms [Fig. 3(c)] in unit of k_BT. First, there is a direct influence from the conjugate of particle number and chemical potential ($\mu\Delta N$); second, there is a change in the potential energy (ΔE); third, the conjugate of available volume ΔV_a, and the volume fraction ϕ ($-\phi\Delta V_a$) [56]. Finally, the contribution in the configurational entropy from remodeling the spatial arrangements ($-\Delta S$) [37,38]. We deduce that divergent subunits contribute more to the energy, entropy, and available volume terms, while convergent subunits contribute primarily through the change in particle number, represented by the $\mu\Delta N$ term.

D. Interactions, abundance, and volume fraction corroborate the emergence of a divergent subunit

With the mechanistic understanding of subunit assembly, we propose to predict whether a subunit is divergent or not from only protein-protein interactions and the abundance—both can be measured by proteome mass spectrometry experimentally. We provided a visual guidance of this prediction with a bidirectional, half-edged graph in Fig. 4(a).

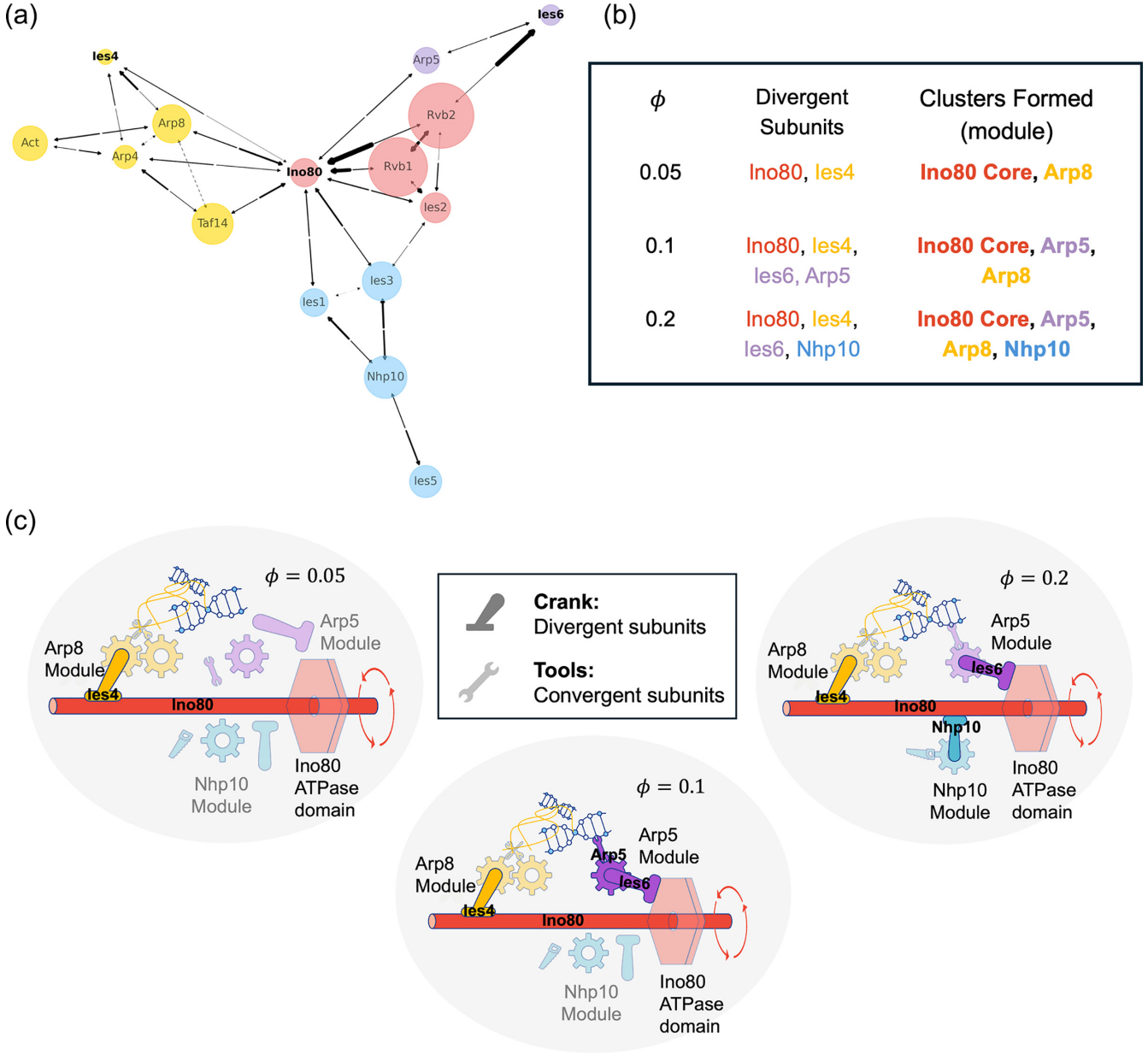

FIG. 4. Divergent subunits orchestrate megacomplexes into operational machinery upon volume fraction changes. (a) A bidirectional, half-edge graphs of interaction networks weighted by subunit abundance for INO80 remodeling complex. Each node corresponds to a subunit, with half-edges indicating interactions with another subunit from the Cross-linking Mass Spectrometry (XL-MS) experiment. Edges are directed to represent the weighted impact from one ("source") to the other ("target"); its width reflects the interaction strength (ϵ_{ij}) weighted by the abundance of the source. The solid edges represent attractive interactions between the source and the target. The only repulsive interactions are represented by a dashed edge between Arp8 and Taf14. Node size reflects subunit abundance, and node color represents modules in the INO80. (b) Divergent subunits and cluster composition at various volume fractions summarized from Fig. S6 in the Supplemental Material [42]. Bold text marks INO80 modules, shown in distinct colors; regular texts indicate subunits, colored according to the modules they belong. (c) A mechanical interpretation of the INO80 remodeling complex in the content of DNA and actin filaments at various volume fractions. Convergent subunits are tools that remodel DNA. Divergent subunits are viewed as cranks that operationally joins the tools together at the correct location along the DNA and actin filaments to perform remodeling functions. The emergence of divergent properties for a subunit depends on volume fraction ($\phi = 0.05,\ 0.1,\ 0.2$). We represent the emergent divergent subunits in opaque, while the convergent ones are transparent.

Because a pair of subunits, i and j, may each have distinct protein abundance ($N_i^{\text{exp}} \neq N_j^{\text{exp}}$), we assigned the interactions as a pair of "source" and "target" into double-headed arrows showing an imbalanced impact to one another due to the many-body effect they each experience. The direction of an arrowhead points from a source subunit i to a target subunit j, and the thickness of the half-edge represents interaction strength weighted by the abundance of the source: $\epsilon_{ij} N_i^{\text{exp}}$.

By observing this graph in Fig. 4(a), we identified the characteristics in the effective potential mean force as well as the uneven abundance between the pair that contribute to the directional impact on a target subunit, leading to its divergence at certain volume fraction (see Table S2 in the Supplemental Material [42]). For example, the number of heterotypic interacting partners is important to the emergence of divergent to Ino80 [bold in Fig. 4(a)]. It has the highest number of partners and is divergent at $\phi = 0.05,\ 0.1$, and 0.2. For another example, low abundance is a contributing factor for Ies4 [bold in Fig. 4(a)]. It has the lowest abundance and is divergent at $\phi = 0.05,\ 0.1$, and 0.2. For Ies6, its divergence is contributed by a combination of relatively strong interaction and relatively low abundance [bold in Fig. 4(a)]. Ies6 shows divergence at $\phi = 0.1$ and 0.2.

In summary, we observed phenomenologically that the divergence of a target subunit satisfies two criteria: (1) when the interaction pointing toward a target subunit j ($I_j = \epsilon_{ij} N_i^{\exp}$) has outweighed its abundance ($N_j^{\exp}$) and (2) when the volume fraction ϕ reaches a certain threshold. We define a ratio of interaction to abundance $\frac{I_j}{N_j^{\exp}}$ for each subunit j. The emergence of a divergent subunit is inferred by a high ratio value in Table S2 in the Supplemental Material [42]. The higher the ratio, the lower the volume fraction is required for a subunit to become divergent. For example, when $\frac{I_j}{N_j^{\exp}} > 40$ (e.g., Ies6 and Ino80), then divergent happens at the volume fraction as low as $\phi = 0.05$. When $\frac{I_j}{N_j^{\exp}} > 30$ (Ies4), the threshold of volume fraction to be divergent increases to $\phi = 0.1$.

IV. DISCUSSION

A. The emergence of a divergent subunit is contributed by system's heterogeneity and many-body effect

The investigation of protein assemblies in crowded, cell-like environments remains challenging. One major difficulty is that the energy landscape governing the assembly of heterotypic subunits in a crowded milieu is highly frustrated [57] due to interactions and abundance, which obscures the underlying physical principles at the cellular level. To reduce this complexity, many existing models assume uniform interactions in a canonical ensemble. For example, Sanders *et al.* [28] modeled stress granule and *P*-body composition by capping subunit binding sites under uniform interaction strengths and fixed protein abundance. While this framework successfully reproduced certain compositional features, it did not explore how variations in interaction heterogeneity or abundance reshape the phase behavior. For the same system, Li and Jacobs [27] introduced a "bridge" particle to study wetting transitions as functions of interaction strength and stoichiometry, again within a canonical ensemble with uniform interactions. Although they solved the phase diagram under these assumptions, alternative cellular states with different protein abundances remain elusive; thus, their theory probed only a narrow region of phase space. As a result, it is difficult to assess how their conclusions extend to more realistic, compositionally diverse assemblies in a changing environment.

Inspired by the finding by Sartori and Leibler [58] that a high volume fraction can drive protein assembly by extending the phase space over interaction energies and chemical potentials (protein abundances). We tackle this problem by using realistic interactions derived from XL-MS and the abundance data from proteomics to determine the boundary of a phase space required to represent protein assemblies *in vivo*. Building on the grand-canonical ensemble approach introduced by Gasic *et al.* [29], we adopt a similar statistical-mechanical framework to a realistic chromatin assembly structure, a system known for its remodeling dynamic of arranging subunits, heterotypic interactions in a crowded, cell-like environment. Within this framework where each subunit is represented by a sphere of equal size, we discovered a surprising "divergent" characteristic of certain subunits, in contrast to others that behave "convergently." This divergence is an emergent property arising from many-body effects in the complex assembly, rather than from any specific pairwise interaction. From a theoretical perspective, this behavior can be rationalized by the concept of spinodal decomposition [55,59]. As illustrated schematically in inset of Fig. 2(a), we consider the free energy as a function of the particle number of a single subunit embedded in the full assembly. For a convergent subunit, its experimental abundance lies within a convex region of the free-energy well, where the corresponding chemical potential—given by the local slope—is well defined and stable. In contrast, for a divergent subunit, the discontinuity and noninvertibility of the abundance-chemical potential relation indicate that its empirical abundance resides within a nonconvex region of the free-energy landscape associated with the transition between dilute and dense phases. In this spinodal-like regime, the grand-canonical ensemble becomes ill-posed, and inference of a unique chemical potential necessarily fails. The observed divergence is therefore a physical signature of thermodynamic instability, not an artifact of the inference procedure.

These convergent and divergent behaviors are not strictly binary but lie on a continuum. In our analysis, Arp5 at a volume fraction of 0.1 and Nhp10 at a volume fraction of 0.2 exhibit critical behavior at the boundary between convergent and divergent regimes. The microscopic origins of these critical phenomena remain to be elucidated and will require further investigation. Nonetheless, by combining realistic structural and abundance data with a grand-canonical description of a crowded, frustrated assembly, our study advances the understanding of general physical principles governing spatial organization of megacomplexes in relation to cell state.

In this work, we primarily aim to capture the first-order physical principles arising from molecular interactions from compositional variations, excluded volumes, and relative abundance, while treating bead-size variation as a higher-order contribution. Keeping the bead size the same allows us to probe frustration and organization in a physiologically relevant megacomplex, thereby bridging a critical gap between idealized theoretical models and experimentally derived structures. This is consistent with the assumptions adopted in other polymer theories for coarse-graining degrees of freedom in a complex system, such as Flory-Huggins polymer theory [60,61]. Nevertheless, variations in bead size, albeit small, are expected to influence both excluded-volume effects and intermolecular energies. For bead types with numerous

interacting partners, the number of interacting partners will grow with bead sizes as they encounter with others spatially. In contrast, for bead types with fewer interacting partners, bead size may contribute primarily through excluded-volume constraints. How these thermodynamic effects may influence a subunit type's propensity to emerge as a divergent subunit would be our future study.

B. Divergent subunits lead to assembly of clusters whose subunit composition and spatial arrangement vary with volume fraction

Here, we focus on the INO80 chromatin-remodeling complex, which is characterized by a dynamic and heterogeneous composition that cannot be captured by a single static structure. INO80 exists in multiple compositional states whose subunit diversity is tightly linked to its functions, including its well-established roles in regulating transcription and cellular metabolism. Despite the importance of this complex, the mechanistic contributions of several key subunits, such as Nhp10, Ies4, and Ies6, have remained unresolved at a structural and thermodynamic level. In our framework, we identify these subunits as divergent: Their behavior changes qualitatively with volume fraction, and their emergent divergence across volume fractions leads to distinct cluster assemblies that differ in subunit composition and spatial organization. Figure 4(b) summarizes the divergent subunits and the clusters they form at each volume fraction (see also Fig. S6 in the Supplemental Material [42]). The resulting visualizations highlight clear morphological differences in assembly architecture across volume fractions, shaped jointly by protein-protein interactions and abundances. We speculate that these emergent clusters form the basis for remodeling INO80 into large megacomplexes that engage DNA and actin filaments in multiple configurations.

Notably, every cluster we observe contains at least one divergent subunit, suggesting that divergence plays a critical role in cluster nucleation and stabilization. Consistent with this, the PMFs in Fig. 2(b) show that divergent subunits act as effective "stickers," and Fig. 3 demonstrates that they strongly modulate global thermodynamic parameters of the assembly, whereas convergent subunits have a comparatively minor impact. We therefore propose that divergent subunits promote nucleation through long-range, many-body interactions that emerge from the collective distribution of subunit interaction strengths and abundances. In this sense, divergent subunits function as "coordinators" of spatial organization, while convergent subunits serve as "participants" whose behavior is largely shaped by the surrounding environment. Although our model does not yet resolve atomic-level details of INO80 assembly, it provides a mechanistic framework for understanding how subunit-specific characteristics and macromolecular volume fraction jointly drive the remodeling dynamics and megacomplex formation of INO80.

C. A machinery view of INO80 remodeling DNAs with divergent subunits that take cues from volume fraction

With a mechanistic view of how volume fraction shapes molecular assemblies by invoking the divergent subunits, we propose a mechanism for INO80, a chromatin remodeling complex (CRC), to anneal the shape and position of the DNA fragment it binds to [Fig. 4(c)]. We liken INO80 to a toolbox, in which each module acts as a "tool" that executes a task. At low volume fractions $\phi = 0.05$, these tools may be loosely connected with each other. We labeled them with transparent colors. As the volume fraction increases to $\phi = 0.1$, some subunits, for example, Arp5 (purple) and Ies6 (purple), become divergent. We labeled them with opaque colors. The Arp5 and Ies6 subunits promote the incorporation of Arp5 module into the assembled machinery into the INO80 CRC complex. In contrast, the state of Nhp10 module and Arp8 module remains unchanged as volume fraction increases from $\phi = 0.05$ to 0.1. As ϕ increases from 0.1 to 0.2, the Nph10 subunit (blue) becomes divergent, and the Nph10 module is further recruited to the machinery of INO80 CRC. Interestingly, although only Ies6 remains divergent, the Arp5 module still remains in the assembled CRC, since it requires only one divergent subunit to assemble.

These observations revealed a set of programmable units that manipulate INO80 CRC, whose biological function is to remodel the position of a DNA fragment via its ATPase domain. We demonstrated that the inner workings of INO80 subunits are influenced by relative abundance, multivalent interactions, and the volume fraction of the system they emerge. Among the known subunits of INO80, the structure of Nhp10 module is intrinsically ordered and unknown. We speculate that its position as a peripheral module could act as a flexible binding target to be recruited by other CRC complexes to form even larger megacomplexes through a "fly-casting mechanism [62–64]".

This perspective considers volume fraction as an active regulatory mechanism for organizing higher-order complex formation. The intracellular environment is inherently crowded [65,66], and macromolecular concentrations significantly influence collective behavior through excluded-volume effects [50,66]. Cells are believed to maintain crowding homeostasis to ensure proper function [39,40,67,68]. We kept the terminology of "macromolecular crowding" to its stringent definition by Rivas and Minton [69]. Our approach expands on this view by treating volume fraction and available volume as thermodynamic conjugate variables [56], similar to the relationship between pressure and volume in mechanical work, with volume fraction effectively proportional to hydrostatic pressure. Unlike active processes that require Adenosine Triphosphate (ATP) and are limited by energy supply, volume fraction can spontaneously influence megacomplex formation in response to physical parameters such as osmotic pressure, cell volume, and macromolecular concentration [39,40]. In this framework, divergent subunits may act as molecular points of crosstalk that sense these environmental cues and initiate higher-order spatial organization of INO80 at the appropriate time and place.

ACKNOWLEDGMENTS

J.W., J.N., and J.H. thanked the support from the National Science Foundation MCB 2221824. M.S.C. thanked the support by the BRaVE project funded by the U. S. Department of Energy (DOE), Office of Science, Office of Biological and Environmental Research, under FWP 81832. Pacific Northwest

National Laboratory is a multi-program national laboratory operated by Battelle for the DOE under Contract DE-AC05-76RL01830.

DATA AVAILABILITY

The data that support the findings of this article are openly available [70].

[1] C. Bunne *et al.*, How to build the virtual cell with artificial intelligence: Priorities and opportunities, Cell **187**, 7045 (2024).

[2] C. G. Johnson *et al.*, Multi-omics reveals temporal scales of carbon metabolism in *Synechococcus elongatus* PCC 7942 under light disturbance, PRX Life **3**, 033017 (2025).

[3] A. Emad and S. Sinha, Inference of phenotype-relevant transcriptional regulatory networks elucidates cancer type-specific regulatory mechanisms in a pan-cancer study, NPJ Syst. Biol. Appl. **7**, 9 (2021).

[4] I. Papatheodorou, A. Oellrich, and D. Smedley, Linking gene expression to phenotypes via pathway information, J. Biomed. Semant. **6**, 17 (2015).

[5] J. Ding, N. Sharon, and Z. Bar-Joseph, Temporal modelling using single-cell transcriptomics, Nat. Rev. Genet. **23**, 355 (2022).

[6] J. Jumper *et al.*, Highly accurate protein structure prediction with AlphaFold, Nature (London) **596**, 583 (2021).

[7] C. Buccitelli and M. Selbach, mRNAs, proteins and the emerging principles of gene expression control, Nat. Rev. Genet. **21**, 630 (2020).

[8] M. R. Spratt and K. Lane, Navigating environmental transitions: The role of phenotypic variation in bacterial responses, mBio **13**, e02212 (2022).

[9] A. Burga and B. Lehner, Beyond genotype to phenotype: Why the phenotype of an individual cannot always be predicted from their genome sequence and the environment that they experience, FEBS J. **279**, 3765 (2012).

[10] T. Misteli, The concept of self-organization in cellular architecture, J. Cell Biol. **155**, 181 (2001).

[11] K. Henzler-Wildman and D. Kern, Dynamic personalities of proteins, Nature (London) **450**, 964 (2007).

[12] S. F. Banani *et al.*, Biomolecular condensates: Organizers of cellular biochemistry, Nat. Rev. Mol. Cell Biol. **18**, 285 (2017).

[13] Y. Shin and C. P. Brangwynne, Liquid phase condensation in cell physiology and disease, Science **357**, eaaf4382 (2017).

[14] C. P. Brangwynne *et al.*, Germline P granules are liquid droplets that localize by controlled dissolution/condensation, Science **324**, 1729 (2009).

[15] P. Li *et al.*, Phase transitions in the assembly of multivalent signalling proteins, Nature (London) **483**, 336 (2012).

[16] A. A. Hyman, C. A. Weber, and F. Jülicher, Liquid-liquid phase separation in biology, Annu. Rev. Cell Dev. Biol. **30**, 39 (2014).

[17] K. Qin, A. R. Fernie, and Y. Zhang, The assembly of supercomplexes in the plant chloroplast, Biomolecules **11**, 1839 (2021).

[18] R. Guo *et al.*, Architecture of human mitochondrial respiratory megacomplex I2III2IV2, Cell **170**, 1247 (2017).

[19] J. Minagawa, State transitions—The molecular remodeling of photosynthetic supercomplexes that controls energy flow in the chloroplast, Biochim. Biophys. Acta **1807**, 897 (2011).

[20] H. Schägger and K. Pfeiffer, Supercomplexes in the respiratory chains of yeast and mammalian mitochondria, EMBO J. **19**, 1777 (2000).

[21] J. D. Forman-Kay *et al.*, What are the distinguishing features and size requirements of biomolecular condensates and their implications for RNA-containing condensates?, RNA **28**, 36 (2022).

[22] A. S. Lyon, W. B. Peeples, and M. K. Rosen, A framework for understanding the functions of biomolecular condensates across scales, Nat. Rev. Mol. Cell Biol. **22**, 215 (2021).

[23] J. Wang *et al.*, A molecular grammar governing the driving forces for phase separation of prion-like RNA binding proteins, Cell **174**, 688 (2018).

[24] A. S. Holehouse and R. V. Pappu, Functional implications of intracellular phase transitions, Biochemistry **57**, 2415 (2018).

[25] S. F. Banani *et al.*, Compositional control of phase-separated cellular bodies, Cell **166**, 651 (2016).

[26] P. Yang *et al.*, G3BP1 is a tunable switch that triggers phase separation to assemble stress granules, Cell **181**, 325 (2020).

[27] T. Li and W. M. Jacobs, Predicting the morphology of multiphase biomolecular condensates from protein interaction networks, PRX Life **2**, 023013 (2024).

[28] D. W. Sanders *et al.*, Competing protein-RNA interaction networks control multiphase intracellular organization, Cell **181**, 306 (2020).

[29] A. G. Gasic, A. Sarkar, and M. S. Cheung, Understanding protein-complex assembly through grand canonical maximum entropy modeling, Phys. Rev. Res. **3**, 033220 (2021).

[30] S. Shukla *et al.*, Conformational switching of Arp5 subunit regulates INO80 chromatin remodeling, Nucl. Ac. Res. **53**, gkae1187 (2025).

[31] H. Wu *et al.*, Reorientation of INO80 on hexasomes reveals basis for mechanistic versatility, Science **381**, 319 (2023).

[32] F. Kunert *et al.*, Structural mechanism of extranucleosomal DNA readout by the INO80 complex, Sci. Adv. **8**, eadd3189 (2022).

[33] X. Zhang *et al.*, Structure and functional interactions of INO80 actin/Arp module, J. Mol. Cell Biol. **11**, 345 (2019).

[34] M. Zhang *et al.*, Hexasome-INO80 complex reveals structural basis of noncanonical nucleosome remodeling, Science **381**, 313 (2023).

[35] A. Tosi *et al.*, Structure and subunit topology of the INO80 chromatin remodeler and its nucleosome complex, Cell **154**, 1207 (2013).

[36] G. J. Gowans *et al.*, INO80 chromatin remodeling coordinates metabolic homeostasis with cell division, Cell Rep. **22**, 611 (2018).

[37] A. P. Minton, Influence of macromolecular crowding upon the stability and state of association of proteins: Predictions and observations, J. Pharm. Sci. **94**, 1668 (2005).

[38] R. J. Ellis, Macromolecular crowding: Obvious but underappreciated, Trends Biochem. Sci. **26**, 597 (2001).

[39] J. van den Berg, A. J. Boersma, and B. Poolman, Microorganisms maintain crowding homeostasis, Nat. Rev. Microbiol. **15**, 309 (2017).

[40] L. J. Holt and M. Delarue, Macromolecular crowding: Sensing without a sensor, Curr. Opin. Cell Biol. **85**, 102269 (2023).
[41] S. Asakura and F. Oosawa, Interaction between particles suspended in solutions of macromolecules, J. Polym. Sci. **33**, 183 (1958).
[42] See Supplemental Material at http://link.aps.org/supplemental/10.1103/b8cy-vfy5 for detailed methods and additional results.
[43] O. Willhoft and D. B. Wigley, INO80 and SWR1 complexes: The non-identical twins of chromatin remodelling, Curr. Opin. Struct. Biol. **61**, 50 (2020).
[44] K. R. Knoll *et al.*, The nuclear actin-containing Arp8 module is a linker DNA sensor driving INO80 chromatin remodeling, Nat. Struct. Mol. Biol. **25**, 823 (2018).
[45] J. Garcia, S. Paul, S. Shukla, Y. Zhong, K. Beauchemin, and B. Bartholomew, Conformational switching of Arp5 subunit differentially regulates INO80 chromatin remodeling, bioRxiv (2024).
[46] P. Bansal, S. Lahiri, C. N. Kumar, *et al.*, Dbf4-dependent kinase finetunes Ino80 function at chromosome replication origins, Nat. Commun. **17**, 3029 (2026).
[47] M. E. Sardiu *et al.*, Conserved abundance and topological features in chromatin-remodeling protein interaction networks, EMBO Rep. **16**, 116 (2015).
[48] B. Zhang and P. G. Wolynes, Topology, structures, and energy landscapes of human chromosomes, Proc. Natl. Acad. Sci. USA **112**, 6062 (2015).
[49] F. Morcos *et al.*, Direct-coupling analysis of residue coevolution captures native contacts across many protein families, Proc. Natl. Acad. Sci. USA **108**, E1293 (2011).
[50] A. P. Minton, Excluded volume as a determinant of macromolecular structure and reactivity, Biopolymers **20**, 2093 (1981).
[51] A. Minton and J. Wilf, Effect of macromolecular crowding upon the structure and function of an enzyme: Glyceraldehyde-3-phosphate dehydrogenase, Biochemistry **20**, 4821 (1981).
[52] A. Stukowski, Visualization and analysis of atomistic simulation data with OVITO–The Open Visualization Tool, Modell. Simul. Mater. Sci. Eng. **18**, 015012 (2010).
[53] A. B. Fulton, How crowded is the cytoplasm?, Cell **30**, 345 (1982).
[54] M. S. Cheung, D. Klimov, and D. Thirumalai, Molecular crowding enhances native state stability and refolding rates, Proc. Natl. Acad. Sci. USA **102**, 4753 (2005).
[55] V. P. Skripov and A. V. Skripov, Spinodal decomposition (phase transitions via unstablestates), Sov. Phys. Usp. **22**, 389 (1979).
[56] A. G. Gasic *et al.*, Critical phenomena in the temperature-pressure-crowding phase diagram of a protein, Phys. Rev. X **9**, 041035 (2019).
[57] W. M. Jacobs and D. Frenkel, Phase transitions in biological systems with many components, Biophys. J. **112**, 683 (2017).
[58] P. Sartori and S. Leibler, Lessons from equilibrium statistical physics regarding the assembly of protein complexes, Proc. Natl. Acad. Sci. USA **117**, 114 (2020).
[59] A. Vainchtein *et al.*, The role of the spinodal region in one-dimensional martensitic phase transitions, Physica D **115**, 29 (1998).
[60] M. L. Huggins, Theory of solutions of high polymers1, J. Am. Chem. Soc. **64**, 1712 (1942).
[61] P. J. Flory, Thermodynamics of high polymer solutions, J. Chem. Phys. **10**, 51 (1942).
[62] B. A. Shoemaker, J. J. Portman, and P. G. Wolynes, Speeding molecular recognition by using the folding funnel: The fly-casting mechanism, Proc. Natl. Acad. Sci. USA **97**, 8868 (2000).
[63] Y. Levy, J. N. Onuchic, and P. G. Wolynes, Fly-casting in protein—DNA binding: Frustration between protein folding and electrostatics facilitates target recognition, J. Am. Chem. Soc. **129**, 738 (2007).
[64] E. Trizac, Y. Levy, and P. G. Wolynes, Capillarity theory for the fly-casting mechanism, Proc. Natl. Acad. Sci. USA **107**, 2746 (2010).
[65] R. J. Ellis and A. P. Minton, Join the crowd, Nature (London) **425**, 27 (2003).
[66] S. B. Zimmerman and S. O. Trach, Estimation of macromolecule concentrations and excluded volume effects for the cytoplasm of Escherichia coli, J. Mol. Biol. **222**, 599 (1991).
[67] A. R. Subramanya and C. R. Boyd-Shiwarski, Molecular crowding: Physiologic sensing and control, Annu. Rev. Physiol. **86**, 429 (2024).
[68] C. Alfano *et al.*, Molecular crowding: The history and development of a scientific paradigm, Chem. Rev. **124**, 3186 (2024).
[69] G. Rivas and A. P. Minton, Macromolecular crowding in vitro, in vivo, and in between, Trends Biochem. Sci **41**, 970 (2016).
[70] J. Wang, J. Nde, A. G. Gasic, J. Haseley, and M. S. Cheung, Simulation code for "Physical principles of building protein megacomplexes in a crowded milieu", GitHub, 2026, https://github.com/Cheung-group/Assembly_Regulator/tree/main.